Regularization and redatuming using least squares and conjugate gradients

Daniel R. Smith, Mrinal K. Sen, University of Texas at Austin, and Robert J. Ferguson, CREWES, University of Calgary

SUMMARY

Irregular spacing of sources and receivers, and dead traces plus noise result in incomplete data. Moreover, phase distortion from a complex near-surface can cause lateral reflector discontinuity that statics cannot handle. As a remedy, we have developed a method to handle irregular data and near-surface complexity as one inversion problem. In particular, we use conjugate gradients for optimization where a weighted, damped least squares approach is used to downward continue data. Data that correspond to the new wavefield at depth are generated by minimization of the residual between the given wavefield and the estimated wavefield. The required extrapolation operators are implemented as spatially variable phase-shifts applied within a Fourier integral operator. The resultant Hessian is extremely costly to compute, so we use the method of conjugate gradients (CG) to avoid direct computation of the Hessian. Our CG approach reduces the total number of operations from $\mathcal{O}\left(n^3\right)$ for direct computation of the Hessian, to $\mathcal{O}\left(n^2\right)$ for the CG method, where n is the number of trace locations.

We use a synthetic data example plus a real data example to demonstrate our simultaneous inversion. The synthetic data are the result of an exploding reflector model where the traces are generated by finite differences. Our setup simulates an irregular, horizontal recording aperture above a line source in which a series of point sources are embedded - the design of the sources results in a flat reflection event and a series of steep diffractions with conflicting dips. The near-surface correction aspect of our CG inversion removes the lateral velocity effects in the synthetic data, and the trace interpolation aspect reconstructs the missing traces.

Our real data example was acquired by Husky Oil Ltd. in the Alberta Foothills of the Canadian Rocky Mountains. Shot spacing for these data is very irregular, and common receiver gathers suffer from incomplete trace coverage as a result. Further, the near-surface is highly heterogeneous due to significant topographic variation and lateral velocity variation, and reflector continuity is compromised as a result. CG inversion of these data successfully reconstructs the data, with some remaining artifacts due to aliasing, and lateral continuity of the reflectors is improved. As a side benefit, because our extrapolation operator is implemented in the temporal and spatial Fourier domain, ground roll is suppressed.

In all instances, we find that the efficiency of the method is improved by an order of magnitude when compared to direct inversion.

LEAST SQUARES INVERSION

For the problem of irregular trace spacing, the square of the l_2 norm $\mathbf{e}$ is [Ferguson (2006), see also Kühl and Sacchi (2003)]

$$E_d = \parallel \mathbf{e} \parallel^2 = \parallel \mathbf{W}_e \left[\psi_z - \mathbf{U}_{-\Delta z} \psi_{z+\Delta z} \right] \parallel^2, \tag{1}$$

where $\mathbf{W}_e$ is a diagonal matrix (non-zero elements on the main diagonal only) that gives unit weight to live traces and zero weight to null traces. (In our approach, the desired, regular trace-spacing is achieved through insertion of null traces.) Alternatively, $\mathbf{W}_e$ can be computed as the inverse of the covariance matrix. The wavefield ψ_z is known at depth z, and $\psi_{z+\Delta z}$ is the desired wavefield at reference depth $z + \Delta z$. The operator $\mathbf{U}_{-\Delta z}$ is a wavefield extrapolation operator that carries wavefields $-\Delta z$ through the medium.

The number of traces estimated in this problem exceeds the number of actual traces, so this problem is underdetermined. Damping, therefore, is required to establish solution uniqueness. Here, damping takes the form of a model norm according to

$$E_m = \parallel \mathbf{W}_m \, \psi_{z+\Delta z} \parallel, \tag{2}$$

where $\mathbf{W}_m$ is a second order spatial derivative used to select the smoothest model. The total cost function E to be minimized, then, is

$$E = E_d + \varepsilon^2 E_m, \tag{3}$$

where ε is a user-defined scalar that is determined by trial and error. The minimum of cost function E with respect to $\psi_{z+\Delta z}$ is (Ferguson, 2006)

$$\psi_{z+\Delta z} = \left[\mathbf{U}^A_{-\Delta z} \mathbf{W}_e \mathbf{U}_{-\Delta z} + \epsilon^2 \mathbf{W}_m \right]^{-1} \mathbf{U}_{-\Delta z} \mathbf{W}_e \psi. \tag{4}$$

EXTRAPOLATION OPERATORS

The required extrapolation operators $U^A_{-\Delta z}$ and $U_{-\Delta z}$ (Margrave and Ferguson, 1999) in our approach are nonstationary in lateral velocity and monochromatic according to

$$\begin{aligned} &\psi \left(x, y, z - \Delta z, \omega \right) = \left[\mathbf{U}_{-\Delta z} \psi_z \right] \left(x, y, z - \Delta z, \omega \right) \\ &= \frac{1}{\left(2\pi \right)^2} \int \varphi \left(k_x, k_y, z, \omega \right) e^{-i \, \Delta z \sqrt{\left(\frac{\omega}{v(x,y)} \right)^2 - k_x^2 - k_y^2}} \\ &e^{-i \left[k_x \, x + k_y \, y \right]} dk_x \, dk_y, \end{aligned} \tag{5}$$

where the x, y, and ω are spatial and temporal frequency coordinates, $\varphi \left(k_x, k_y, z, \omega \right)$ is the Fourier spectrum of wavefield ψ_z (limits $-\infty$ to $+\infty$ are omitted for brevity), and k_x and k_y are wavenumbers that correspond to x and y respectively. $\mathbf{U}^A_{-\Delta z}$ is simply the complex conjugate of $\mathbf{U}_{-\Delta z}$. The $\omega \rightarrow t$ transform completes extrapolation. Manipulation of the sign on Δz in the complex exponential in equation 5 ensures exponential decay of the evanescent region of k_x, ω space.

CONJUGATE GRADIENTS

Because $\mathbf{U}^{A}_{-\Delta z}\mathbf{W}_e\mathbf{U}_{-\Delta z}$ is a positive definite matrix, the solution is the minimum of a quadratic form, and the residuals are the direction of steepest descent and are orthogonal to each other. Based on the method of conjugate gradients (CG), a solution to the minimization problem is found through a search in the direction of the conjugation of these residuals. The true residuals are $\mathbf{r} = \psi_z - \mathbf{U}_{-\Delta z}\psi_{z+\Delta z}$ whereas in the conjugate gradient method the residual polynomials are constructed. A line search is then conducted to find the step length γ by minimizing the quadratic form in a straight line along the search direction. If the start point is $\mathbf{x}_o$, which can be completely arbitrary, each update thereafter to the model is

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \gamma_i \mathbf{p}_i, \tag{6}$$

where $\mathbf{p}_o = \mathbf{r}_o$. The step length γ is

$$\gamma_i = \frac{\mathbf{r}^t \mathbf{r}_i}{\mathbf{p}_i^t \mathbf{A} \mathbf{p}_i} \tag{7}$$

as derived by Shewchuk (1994), and $\mathbf{A}$ is the augmented matrix $\begin{bmatrix} \mathbf{W}_e\mathbf{U}_{-\Delta z} \\ \epsilon \mathbf{W}_m \end{bmatrix}$. The new residual is constructed as

$$\mathbf{r}_{i+1} = \mathbf{r}_i - \gamma \mathbf{A} \mathbf{p}_i. \tag{8}$$

The new direction is made as a conjugation of this new residual being

$$\mathbf{p}_{i+1} = \mathbf{r}_{i+1} + \beta_{i+1}\mathbf{p}_i, \tag{9}$$

where β_{i+1} is a scalar multiplier derived from the conjugate Gram-Schmidt process. It is defined as

$$\beta_{i+1} = \frac{\mathbf{r}_{i+1}\mathbf{r}_{i+1}}{\mathbf{r}_i^t \mathbf{r}_i}, \tag{10}$$

taken from Shewchuk (1994). The final model update $\mathbf{x}_k$ is the regularized and redatumed spectrum for a single frequency $\psi_{z+\Delta z}$. The solution should converge in no more than n iterations. If the eigenvalues of the operators are spread over a small scale, or there are clusters of eigenvalues, or even eigenvalues with a multiplicity larger than one, then the solution should converge in $< n$ iterations. The method can be made to terminate at a given tolerance, and the final model is the solution.

SYNTHETIC EXAMPLE

A synthetic example is derived to apply the method of conjugate gradients in order to test the viability of this method for regularization and datuming. The velocity model is laterally heterogeneous and increasing linearly. The model setup (Figure 1) is that of an exploding reflector model with a flat line of weak sources and five strong point sources at a depth of 300m. The receiver array is an irregularly spaced receiver to simulate a highly irregular data set and is located at a depth of 100m. This was done to demonstrate the stability of the method over large extrapolation distances. The data generated by this model is shown in Figure 2a. A 1D Fourier transform from $t - \omega$ was applied to the data, and each spectrum for a certain frequency was then fed into a conjugate gradient algorithm. A

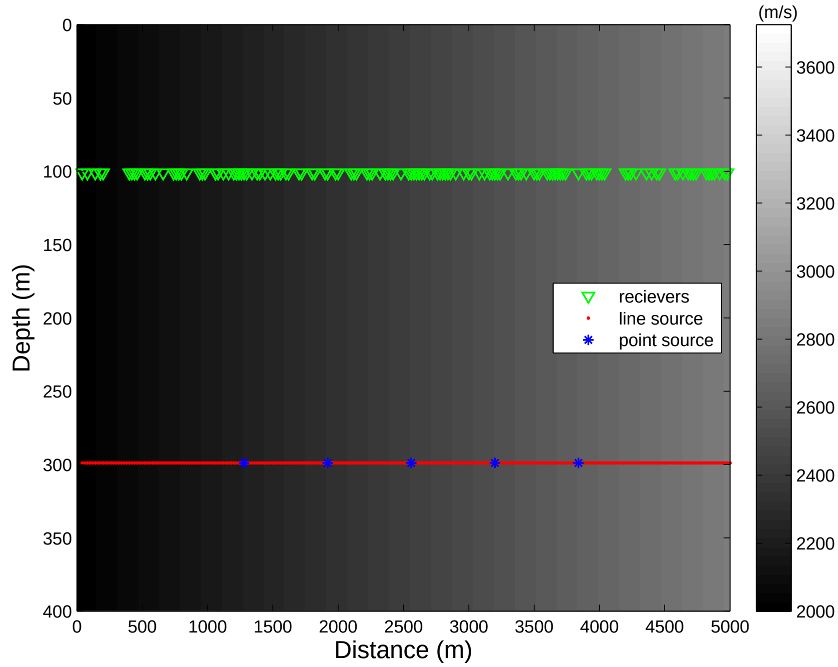

FIG. 1. The model setup used in the synthetic example. It simulates a line source (diamonds) and five point sources (large diamonds) and a highly irregular receiver array (triangles), (Ferguson, 2006).

pseudo-code is displayed in table 1. An isotropic phase-shift extrapolation operator was then applied to the solution to back-extrapolate the wavefield. A band-limited 1D inverse Fourier transform was applied to get the results displayed in Figure 2b. The smoothing parameter ϵ is found by trial and error, and a value of $\epsilon = 0.5$ was found to produce the desired results, reducing the noise sufficiently but not oversmoothing the result. The tolerance level was 1%. The direct Newton's method solution is also displayed for comparison in Figure 2c. Due to the computation of the Hessian within the least squares solution, the direct Newton is costly. An approximation can be applied to make the Hessian more efficient as done by Ferguson (2006), but this imposes a dip limitation onto the outcome. The computations for one iteration using Newton's method are $\mathcal{O}(n^3)$. For a complete 3D survey the computation is $\mathcal{O}(10^{17})$ operations. The computational cost for the conjugate gradient method for each iteration is $\mathcal{O}(n^2)$ which translates into a computational cost for a 3D survey to be $\mathcal{O}(10^{15})$ operations. The data regularized with the conjugate gradients successfully filled in the missing amplitudes and removed the velocity effects by flattening the line and point sources. Compared with Newton's method, there is no noticeable difference.

CANADIAN FOOTHILLS EXAMPLE

The real data example is from the Canadian foothills in Alberta. It was acquired by Husky Oil Ltd., and it was recorded in an overthrust belt region. The region is characterized by overthrust structures of various geometric complexity and stratigraphy units ranging from carbonates, shales as well as other clastics, (Stork et al., 1995). The velocity model for this area was derived by turning-wave tomography (Figure 3) as the near-surface

Table 1. The Least-Squares Conjugate Gradient algorithm following van den Eshof and Sleijpen (2004).

$$
\begin{aligned}
&\mathbf{s}_o = \mathbf{b}; \mathbf{r}_o = \mathbf{A}^T\mathbf{s}_o \\
&\mathbf{p}_o = \mathbf{r}_o; \mathbf{x}_o = 0 \\
&\text{for } k = 1 : k_{max} \\
&\quad \mathbf{q}_{k-1} = \mathbf{A}\mathbf{p}_{k-1} \\
&\quad \gamma_{k-1} = \frac{\|\mathbf{r}_{k-1}\|^2}{\|\mathbf{q}_{k-1}\|^2 + \epsilon\|\mathbf{p}_{k-1}\|^2} \\
&\quad \mathbf{x}_k = \mathbf{x}_{k-1} + \gamma_{k-1}\mathbf{p}_{k-1} \\
&\quad \mathbf{s}_k = \mathbf{s}_{k-1} - \gamma_{k-1}\mathbf{q}_{k-1} \\
&\quad \mathbf{r}_k = \mathbf{A}^T\mathbf{s}_k - \gamma_{k-1}\mathbf{q}_{k-1} \\
&\quad \beta = \frac{\|\mathbf{r}_k\|^2}{\|\mathbf{r}_{k-1}\|^2} \\
&\quad \mathbf{p}_k = \mathbf{r}_k + \beta_{k-1}\mathbf{p}_{k-1} \\
&\text{end} \\
&\psi_{z+\Delta z} = \mathbf{x}_k
\end{aligned}
$$

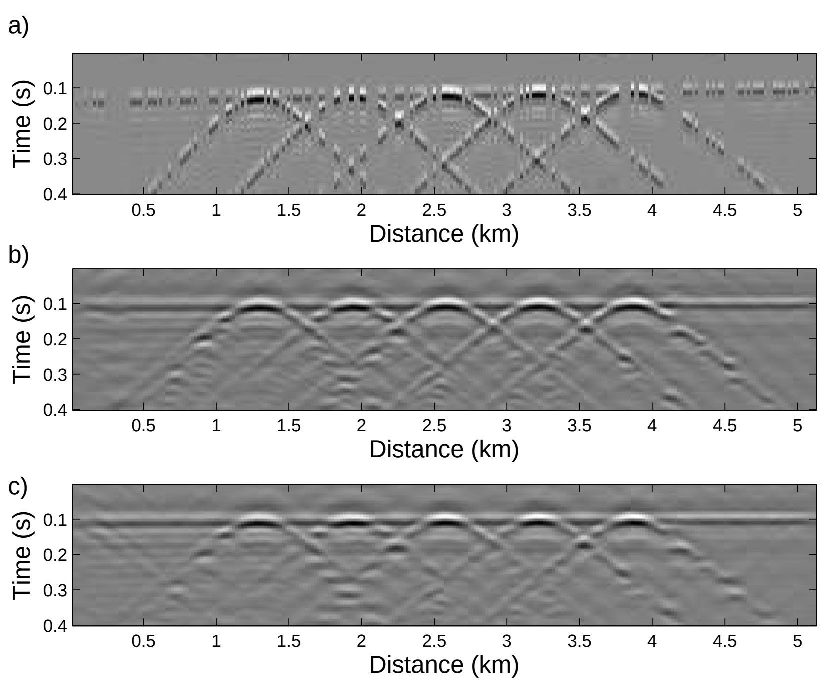

FIG. 2. A view of the results for the synthetic seismic data set. a) The input data. b) the solution using the conjugate gradient scheme and c) Newton's least-square solution.

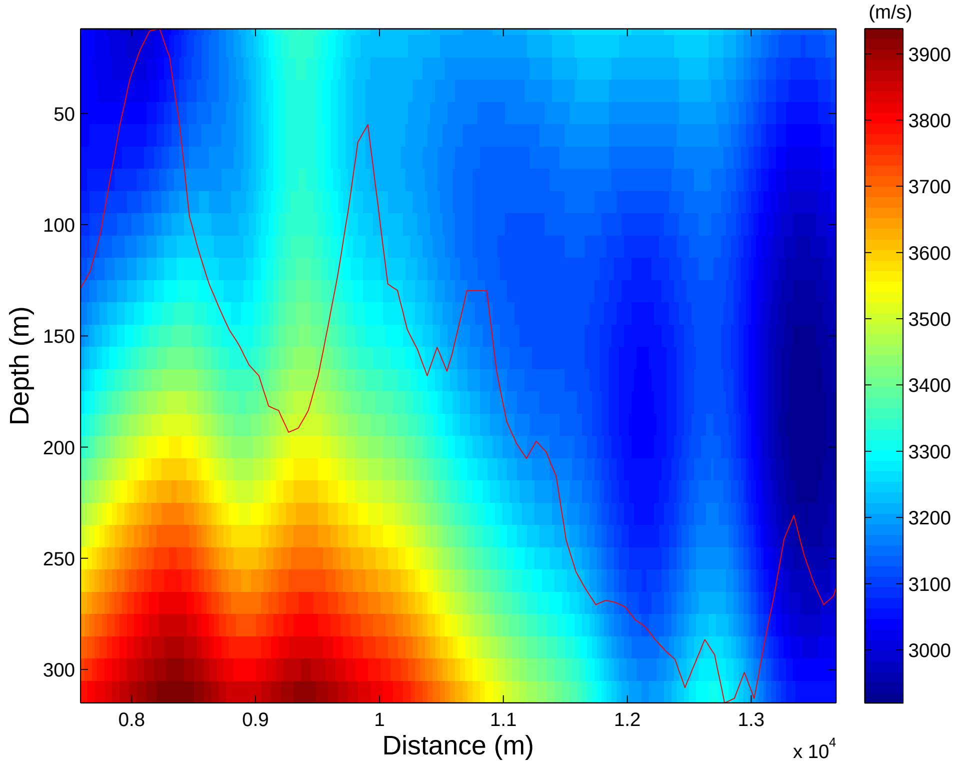

FIG. 3. Velocity model used to regularize and redatum the shot record. The topography is shown in red.

heterogeneity did not allow for refraction velocity analysis.

A common source gather (Figure 4a) was used to demonstrate the effectiveness of the method described in this paper. The receiver array is regularly spaced for the most part so the gather was decimated (Figure 5a) from 300 traces to 56 traces. The conjugate gradient based method was applied to the two gathers following Reshef (1991) and using the velocities in Figure 3. The redatumed results in Figure 4b show improved lateral continuity in the reflections. The regularization and redatumed gather in Figure 5b shows the robustness of the method in that it successfully reconstructs the data and improves the lateral continuity of the reflectors.

CONCLUSIONS

A method for regularization and redatuming is presented using conjugate gradients for the optimization of the solution for weighted, damped least squares. The method is used to downward continue irregular data onto a regularly spaced grid. This is necessary for future processing steps such as fast Fourier transforms and imaging techniques which are susceptible to spatial aliasing. This method is applied to highly irregularly sampled synthetic data as well as real data to demonstrate its effectiveness.

Conjugate gradients are employed to improve the efficiency of computing the Hessian by reducing the computation of the Hessian from matrix-matrix multiplication to vector-matrix multiplication. This cost is reduced by 10 in 2D and we expect to see greater cost

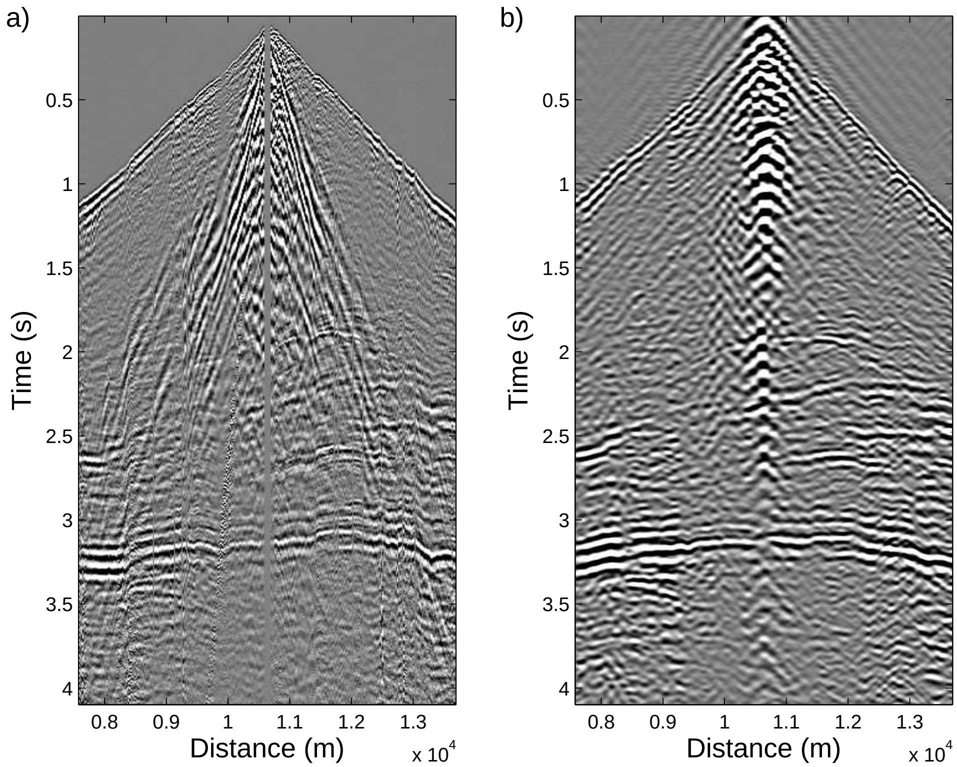

FIG. 4. Input data a) Shot record, b) Redatumed results of the input data in Figure 4a using CG method.

savings in 3D.

This method successfully regularizes and removes the lateral velocity effects of a synthetic model. The line source is continuous and the steep dips are restored. The method also successfully reconstructs the real data with some artifacts and also removes the phase distortions caused by the near-surface which is apparent in the improved lateral continuity of the reflectors.

ACKNOWLEDGMENTS

We thank ConocoPhillips, the Jackson School of Geosciences, University of Texas at Austin, the EDGER Forum, the Institute of Geophysics, and the sponsors of CREWES for supporting this research.

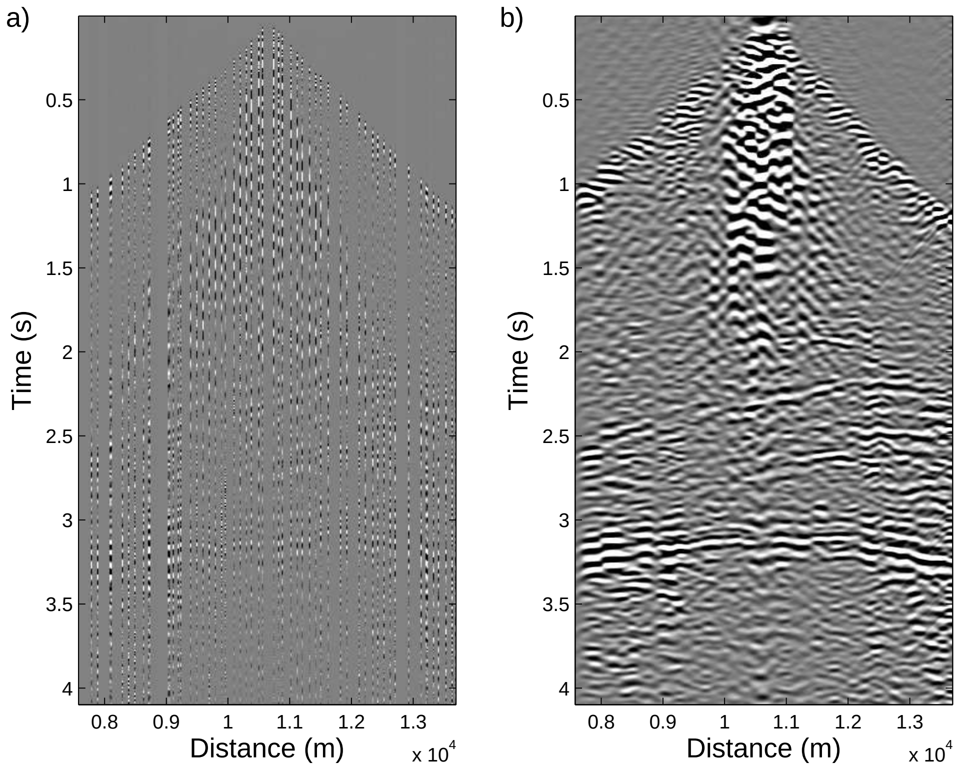

FIG. 5. Input data a) Decimated record, b) Regularized and redatumed results of the input data in Figure 5a using CG method.

REFERENCES

Ferguson, R. J., 2006, Regularization and datuming of seismic data by weighted, damped least squares: Geophysics, **71**, No. 5, U67–U76.

Kühl, H., and Sacchi, M. D., 2003, Least-squares wave-equation migration for avp/ava inversion: Geophysics, **68**, No. 1, 262–273.

Margrave, G. F., and Ferguson, R. J., 1999, Wavefield extrapolation by nonstationary phase shift: Geophysics, **64**, No. 4, 1067–1078.

Reshef, M., 1991, Depth migration from irregular surfaces with depth extrapolation methods: Geophysics, **56**, No. 1, 119–122.

Shewchuk, J. R., 1994, An introduction to the conjugate gradient method without the agonizing pain, Technical paper, School of Computer Science, Carnegie Mellon University, $1\frac{1}{4}$ edition.

Stork, C., Welsh, C., and Skuce, A., 1995, Demonstration of processing and model building methods on a real complex structure data set, Society of Exploration Geophysicists.

van den Eshof, J., and Sleijpen, G. L. G., 2004, Accurate conjugate gradient methods for families of shifted systems: Applied Numerical Mathematics, **49**, 17–37.